10/541118

JC20 Rec'd PCT/PTO 30 JUN 2005

**Method and device for the detection of points of interest in a digital source, corresponding computer program and data carrier**

1. Field of the invention

The field of the invention is that of the detection of points of interest, also called salient points, in a digital image. More specifically, the invention relates to a technique for the detection of points of interest implementing a wavelet-type approach.

A point of interest may be considered to be the representative of a spatial region of the image conveying a substantial portion of information.

Historically, the notion of the salient point has been proposed in the field of computer vision, where one of the major problems consists of the detection of the corners of the objects (whence the term "salient" used here below as a synonym for the term "of interest"). This term was subsequently broadened to include other characteristics of images such as contours, junctions etc.

In image processing, the detection of the salient points corresponding to the corners of the objects is of little interest. Indeed, the corners are generally isolated points, representing only a small part of the information contained in the image. Furthermore, their detection generates heaps of salient points in the textured or noisy regions.

Various other techniques have been proposed, relating especially to the salient points corresponding to the high frequency zones, namely to the contours of the objects. The invention can be applied more specifically to this type of technique.

A more detailed description is given here below of the different techniques for the detection of salient points.

2. Prior art

The detection of salient points (also called points of interest) in images is a problem that has given rise to much research for many years. This section presents the main approaches classically used in the literature. Reference may be made to the document [5] (the documents referred to are listed together in the

appendix B) for a more detailed review of the prior art.

One of the first methods was proposed by Harris and Stephens [7] for the detection of corners. Points of this type were deemed then to convey a major quantity of information and were applied in the field of computer vision.

To define this detector, the following quantity is defined at each point *p(x,y)* of the image $I$:

$$R_{x,y} = Det(M_{x,y}) - kTr(M_{x,y})^2$$

where $M_{x,y}$ is a matrix defined by:

$$M_{x,y} = G(\sigma) \otimes \begin{bmatrix} I_x^2(x,y) & I_x(x,y)I_y(x,y) \\ I_x(x,y)I_y(x,y) & I_y^2(x,y) \end{bmatrix}$$

where:

- $G(\sigma)$ denotes a Gaussian kernel with variance $\sigma^2$ ;
- $\otimes$ denotes the convolution product;
- $I_x$ (resp. $I_y$) denotes the first derivative of $I$ following the direction $x$ (resp. $y$) ;
- $Det(M_{x,y})$ denotes the determinant of the matrix $M_{x,y}$ ;
- $Tr(M_{x,y})$ denotes the trace of the matrix $M_{x,y}$ ;
- $k$ is a constant generally used with a value of 0.04.

The salient points are then defined by the positive local extreme values of the quantity $R_{x,y}$.

In [5], the authors also propose a more precise version of the Harris and Stephens detector. This version replaces the computation of the derivatives of the image $I$ by a precise computation of the Gaussian kernel.

The Harris and Stephens detector presented here above has been extended to the case of color pictures in [6]. To do this, the authors extend the definition of the matrix $M_{x,y}$ which then becomes:

$$M_{x,y} = G(\sigma) \otimes \begin{bmatrix} (R_x^2 + G_x^2 + B_x^2)(x,y) & (R_xR_y + G_xG_y + B_xB_y)(x,y) \\ (R_xR_y + G_xG_y + B_xB_y)(x,y) & (R_y^2 + G_y^2 + B_y^2)(x,y) \end{bmatrix}$$

where:

- $R_x, G_x, B_x$ respectively denote the first derivatives of the red, green and blue colorimetrical planes in the direction $x$ ;
- $R_y, G_y, B_y$ respectively denote the first derivatives of the red, green and blue colorimetrical planes in the direction $y$ ;

In [10], the authors consider the salient points to be the points of the image showing high contrast. To build a detector of this kind, the authors use a multiple-resolution approach based on the construction of a Gaussian pyramid.

Let it be assumed that the image $I$ has a size $2^N x 2^N$. We can define a pyramid with $N$ levels where the level 0 corresponds to the original image and the level $N-1$ corresponds to a one-pixel image.

At the level $k$ of the pyramid, the contrast of the point $P$ is defined by:

$$C_k(P)=\frac{G_k(P)}{B_k(P)} \text{ with } 0 \le k \le N-1 \text{ and } C_N(P)=1$$

where $G_k(P)$ defines the local luminance at the point $P$ and at the level $k$, and $B_k(P)$ defines the luminance of the local background at the point $P$ and at the level $k$.

These two variables are computed at each point and for each level of the pyramid. They can therefore be represented by two pyramids called a luminance pyramid and a background pyramid and defined by:

$$G_k(P) = \sum_{M \in Fils(P)} w(M) G_{k-1}(M)$$

$$B_k(P) = \sum_{Q \in Parent(P)} W(Q) G_{k+1}(Q)$$

where:

- The notations *Offspring (P)* and *Parent(P)* denote the hierarchical relationships in the Gaussian pyramid;
- $w$ is a standardized weight function that can be adjusted in order to simulate the Gaussian pyramid;
- $W$ is a standardized weight function taking account of the way in which $P$ is used to build a luminance of its ancestors in the pyramid.

In this approach, a salient point is a point characterized by a high value of the local contrast. In order to take account of the non-symmetry of the variable

$C_k$, the authors introduce a new variable in order to obtain a zero value for a situation of non-contrast and a value > 0 everywhere else.

This new variable is defined by:

$$C_k^{\bullet}(P)=Min\left(\frac{|G_k(P)-B_k(P)|}{B_k(P)},\frac{|G_k(P)-B_k(P)|}{255-B_k(P)}\right).$$

With this new variable, the salient points are defined by the local maximum values of $C_k^{\bullet}$ greater than a fixed threshold.

The salient points detector initially presented in [11] is doubtless the closest to the present invention since it is also based on the use of the theory of wavelets. Indeed, it is the view of the authors that the points conveying a major part of the information are localized in the regions of the image having high frequencies.

By using wavelets with compact carriers, the authors are capable of determining the set of points of the signal $f$ (assumed for the time being to be one-dimensional) that were used to compute any wavelet coefficient whatsoever $D_{2^j}f(n)$, and can do so at any resolution whatsoever $2^j\ (j\leq -1)$.

On the basis of this observation, the hierarchy of wavelet coefficients is built. For each resolution level and for each wavelet coefficient $D_{2^j}f(n)$ of this level, this hierarchy determines the set of wavelet coefficients of the immediately higher level of resolution $2^{j+1}$ necessary to compute $D_{2^j}f(n)$:

$$C(D_{2^j}f(n))=\{D_{2^{j+1}}f(k),2n\leq k\leq 2n+2p-1\},0\leq n<2^jN$$

where $p$ denotes the regularity of the wavelet base used (i.e. the size of the wavelet filter) and $N$ denotes the length of the original signal $f$.

Thus, each wavelet coefficient $D_{2^j}f(n)$ is computed from $2^{-j}p$ points of the signal $f$. Its offspring coefficients $C(D_{2^j}f(n))$ give the variation of a subset of these $2^{-j}p$ points. The most salient subset is the one whose wavelet coefficient is the maximum (in absolute value) at the resolution level $2^{j+1}$.

This coefficient therefore needs to be considered at this level of resolution. By applying this process recursively, a coefficient $D_{2^{-1}}f(n)$ is selected with the

resolution $\frac{1}{2}$. This coefficient represents *2p* points of the signal *f*. To select the corresponding salient point in *f*, the authors propose to choose that point, among these *2p* points, whose gradient is the maximum in terms of absolute value.

To extend this approach to the 2D signals constituted by the images, the authors apply the same approach to each of the three subbands $D^1_{2^j}I, D^2_{2^j}I, D^3_{2^j}I$ where *I* denotes the original image. In the case of the images, the spatial carrier of the wavelet base is sized *2p*x*2p*. Thus, the cardinal of $C(D^s_{2^j}f(x,y))$ is $4p^2$ for any $s=1,2,3$. For each orientation (horizontal, vertical and oblique), the method makes a search, among the offspring coefficients of a given coefficient, for the one whose amplitude is the maximum. If different coefficients of different orientations lead to the same pixel of *I*, then this pixel is considered to be a salient point.

This technique has been used especially in image indexation in [9].

### 3. Drawbacks of the Prior Art

As shown in the previous section, many methods have been proposed in the literature for the detection of salient points.

The major difference between these approaches relies on the very definition of a salient point. Historically, researchers in the field of computer vision have devoted attention to the corners of objects. It is thus that the Harris and Stephens detector [7] was proposed. This detector has recently been extended to color in [6]. The corners of objects do not, however, represent any relevant information in the field of image processing. Indeed, in the case of weakly textured images, these dots will be scattered in space and will not give any satisfactory representation of the image. In the case of textured or noisy images, the dots will all be concentrated in the textures and within a local and non-comprehensive representation of the image.

The definition of contrast-based salience [10] is appreciably more interesting for image processing. Unfortunately, this approach suffers from the same defect as the previous one in the case of textured or noisy regions.

❖ The wavelet-based approach proposed by E. Loupias and N. Sebe [11] is

clearly the most robust and most worthwhile approach. Indeed, it has long been known that the contours represent the primary information of an image since it perfectly matches the human visual system.

## 4. Goals and characteristics of the invention

It is therefore a particular aim of the invention to overcome the different drawbacks of the prior art.

More specifically, it is an aim of the invention to provide a technique for the detection of salient points corresponding to a high frequency, and giving preference to no particular direction in the image.

It is another aim of the invention to provide such a technique that calls for a reduced number of operations as compared with prior art techniques.

In particular, it is a goal of the invention to provide a technique of this kind enabling the use of wavelet bases with a large-sized carrier.

These goals, as well as others that shall appear more clearly here below, are achieved by means of a method for the detection of points of interest in a source digital image, said method implementing a wavelet transformation associating a sub-sampled image, called a scale image, with a source image, and wavelet coefficients corresponding to at least one detail image, for at least one level of decomposition, a point of interest being a point associated with a region of the image showing high frequencies.

According to the invention, this method comprises the following steps:

- the application of said wavelet transformation to said source image;
- the construction of a unique tree structure from the wavelet coefficients of each of said detail images;
- the selection of at least one point of interest by analysis of said tree structure.

In the present document, for the sake of simplification, the term "source image" is applied to an original image or an image having undergone pre-processing (gradient computations, change of colorimetrical space etc.).

Advantageously, for each level of decomposition, at least two detail images, respectively corresponding to at least two directions predetermined by said wavelet transformation, are determined.

This wavelet transformation may use especially first-generation or second-generation (mesh-based) wavelets.

In particular, the detail images may comprise:

- a detail image representing the vertical high frequencies;
- a detail image representing the horizontal high frequencies;
- a detail image representing the diagonal high frequencies.

Advantageously, the method of the invention comprises a step for merging the coefficients of said detail images so as not to give preference to any direction of said source image.

Advantageously, said step for the construction of a tree structure relies on a zerotree type of approach.

Thus, preferably, each point of the scale image having minimum resolution is the root of a tree with which is associated at least one offspring node respectively formed by each of the wavelet coefficients of each of said detail image or images localized at the same position, and then recursively, four offspring nodes are associated with each offspring node of a given level of resolution, these four associated offspring nodes being formed by the wavelet coefficients of the detail image that is of a same type and at the previous resolution level, associated with the corresponding region of the source image.

According to an advantageous aspect of the invention, said selection step implements a step for the construction of at least one salience map, assigning said wavelet coefficients a salience value representing its interest. Preferably, a salience map is built for each of said resolution levels.

Advantageously, for each of said salience maps, for each salience value, a merging is performed of the pieces of information associated with the three wavelet coefficients corresponding to the three detail images so as not to give preference to any direction in the image.

According to a preferred aspect of the invention, a salience value of a given wavelet coefficient having a given level of resolution takes account of the salience value or values of the descending-order wavelet coefficients in said tree structure of said given wavelet coefficient.

Preferably, a salience value is a linear relationship of the associated wavelet coefficients.

In a particular embodiment of the invention, the salience value of a given wavelet coefficient is computed from the following equations:

$$\begin{cases} S_{2^{-1}}(x,y) = \alpha_{-1}\left(\frac{1}{3}\sum_{u=1}^{3}\frac{D_{2^{-1}}^{u}(x,y)}{Max(D_{2^{-1}}^{u})}\right) \\ S_{2^{j}}(x,y) = \frac{1}{2}\left(\alpha_{j}\left(\frac{1}{3}\sum_{u=1}^{3}\frac{D_{2^{j}}^{u}(x,y)}{Max(D_{2^{j}}^{u})}\right) + \frac{1}{4}\sum_{u=0}^{1}\sum_{v=0}^{1}S_{2^{j+1}}(2x+u,2y+v)\right) \end{cases}$$

In these equations, the parameter $\alpha_k$ may for example have a value $-1/r$ for all the values of k.

According to another preferred aspect of the invention, said selection step comprises a step for building a tree structure of said salience values, the step advantageously relying on a zerotree type approach.

In this case, said selection step advantageously comprises the steps of:

- descending-order sorting of the salience values of the salience map corresponding to the minimum resolution;
- selection of the branch having the highest salience value for each of the trees thus sorted out.

According to a preferred aspect of the invention, said step for the selection of the branch having the highest salience value implements a corresponding scan of the tree starting from its root and a selection, at each level of the tree, of the offspring node having the highest salience value.

As already mentioned, the invention enables the use of numerous wavelet transformations. One particular embodiment implements the Haar base.

One particular embodiment chooses a minimum level of resolution $2^{-4}$.

The method of the invention may furthermore include a step for the computation of an image signature, from a predetermined number of points of interest of said image.

Said signature may thus be used especially to index images by their content.

More generally, the invention can be applied in many fields, and for example for:

- image watermarking;
- image indexing;
- the detection of faces in an image.

The invention also relates to devices for the detection of points of interest in a source digital image implementing the method as described here above.

The invention also relates to computer programs comprising program code instructions for the execution of the steps of the method for the detection of points of interest described here above, and the carriers of digital data that can be used by a computer carrying such a program.

Other characteristics and advantages of the invention shall appear from the following description of a preferred embodiment, given by way of a simple illustrative and non-exhaustive example and from the appended drawings, of which:

- Figure 1 illustrates the principle of multi-resolution analysis of an image I by wavelet transformation;
- Figure 2 presents a schematic view of a wavelet transformation;
- Figure 3 provides a view of a tree structure of wavelet coefficients according to the invention;
- Figure 4 presents an example of salience maps and of the corresponding salience trees;
- Figure 5 illustrates the salience of a branch of the tree of figure 4 ;
- Figures 6a and 6b illustrate experimental results of the method of the invention, Figure 6a showing two original images and Figure

6b showing the corresponding salient points;

- Figure 7 illustrates an image indexing method implementing the detection method of the invention.

## 5. Identification of the essential technical elements of the invention

### *5.0 General principles*

One aim of the invention therefore is the detection of the salient points of an image *I*. These points correspond to the pixels of *I* belonging to high-frequency regions. This detection is based on wavelet theory [1][2][3]. Appendix A briefly presents this theory.

Wavelet transform is a multi-resolution representation of the image enabling the image to be expressed at the different resolutions $\frac{1}{2},\frac{1}{4},etc.$. Thus, at each level of resolution $2^j\,(j \leq -1)$, the wavelet transform represents the image *I*, sized $n \times m = 2^k \times 2^l\ (k,l \in Z)$, in the form:

- a coarse image $A_{2^j}I$ ;
- a detail image $D^1_{2^j}I$ representing the vertical high frequencies (i.e. the horizontal contours);
- a detail image $D^2_{2^j}I$ representing the horizontal high frequencies (i.e. the vertical contours);
- a detail image $D^3_{2^j}I$ representing the diagonal high frequencies (i.e. the corners).

Each of these images is sized $2^{k+j} \times 2^{l+j}$ . Figure 1 illustrates this type of representation.

Each of these three images is obtained from $A_{2^{j+1}}I$ by a filtering followed by a sub-sampling by a factor of two as shown in figure 2. It must be noted that we have $A_{2^0}I = I$.

The invention therefore consists of choosing first of all a wavelet base and a minimum level of resolution $2^r\,(r \leq -1)$. Once the wavelet transformation has been effected, it is proposed to scan each of the three detail images $D^1_{2^r}I$, $D^2_{2^r}I$ and $D^3_{2^r}I$ in order to build a tree structure of wavelet coefficients. This tree

structure is based on the zerotree approach [4], initially proposed for the image encoding. It enables the positioning of a salience map sized $2^{k+r} \times 2^{l+r}$ reflecting the importance of each wavelet coefficient at the resolution $2^r (r \leq -1)$.

Thus a coefficient having significant salience corresponds to a region of $I$ having high frequencies. Indeed, a wavelet coefficient having a high-value modulus at the resolution $2^r (r \leq -1)$ corresponds to a contour of the image $A_{2^{r+1}} I$ along a particular direction (horizontal, vertical or oblique). The zerotree approach tells us that each of the wavelet coefficients at the resolution $2^r$ corresponds to a spatial zone sized $2^{-r} \times 2^{-r}$ in the image $I$.

From the built-up salience map, the invention proposes a method for the choosing, from among of the $2^{-r} \times 2^{-r}$ pixels of $I$, of the pixel that most represents this zone.

In terms of potential applications, the detection of salient points in the images may be used non-exhaustively for the following operations:

- Image watermarking: in this case, the salient points give information on the possible localization of the mark in order to ensure its robustness;
- Image indexing: in detecting a fixed number of salient points, it possible to deduce a signature of the image from it (based for example on colorimetry around salient points) which may then be used for the computation of inter-image similarities;
- Detection of faces: among the salient points corresponding to the high frequencies of the image, some are localized on the facial characteristics (eyes, nose, mouth) of the faces present in the image. They may then be used in a process of detection of faces in the images.

The technique of the invention differs from that proposed by E. Loupias and N. Sebe [11]. The main differences are:

- The salient point search algorithm proposed by Loupias and Sebe requires a search among $2^{2j} \times 4p^2 \times 3$ coefficients for each level of resolution $2^j$ and for a square image. Our algorithm is independent of the size of the wavelet base carrier, leading to a search from among

$2^{2j} \times 4 \times 3$ coefficients. This advantage enables the use of the wavelet bases with a carrier that may be large-sized while most of the publications using the Loupias and Sebe detector use the Haar base which is far from being optimal.

The Loupias and Sebe method considers the subbands independently of each other thus leading them to the detection, by priority, of the maximum gradient points in every direction (i.e. the corners). For our part, we merge the information contained in the different subbands so that no preference is given to any particular direction.

### *5.1 Wavelet tranformation*

Wavelet transformation is a powerful mathematical tool for the multi-resolution analysis of a function [1][2][3]. Appendix A provides a quick overview of this tool.

In the invention, the functions considered are digital images, i.e. discrete 2D functions. Without overlooking general aspects, we assume here that the processed images are sampled on a discrete grid of $n$ lines and $m$ columns with value range in a sampled luminance space containing 256 values. Furthermore it is assumed that $n = 2^k$ $(k \in Z)$ and that $m = 2^l$ $(l \in Z)$.

If the original image is referenced $I$, we then have :

$$I : \left| \begin{array}{l} [0,m] \times [0,n] \rightarrow [0,255] \\ (x,y) \text{ a } I(x,y) \end{array} \right. .$$

As mentioned in section 4, the wavelet transformation of $I$ enables a multi-resolution representation of $I$. At each level of resolution $2^j (j \leq -1)$, the representation of $I$ is given by a coarse image $A_{2^j} I$ and by three detail images $D^1_{2^j} I$, $D^2_{2^j} I$ and $D^3_{2^j} I$. Each of these images is sized $2^{k+j} \times 2^{l+j}$. This process is illustrated in figure 2.

Wavelet transformation necessitates the choice of a scale function $\Phi(x)$ as well as the choice of a wavelet function $\Psi(x)$. From these two functions, a scale filter $H$ and a wavelet filter $G$ are derived, their respective pulse responses $h$ and $g$ being defined by :

$$h(n) = \langle \phi_{2^{-1}}(u), \phi(u-n) \rangle \forall n \in Z$$
$$g(n) = \langle \psi_{2^{-1}}(u), \phi(u-n) \rangle \forall n \in Z.$$

Let $\tilde{H}$ and $\tilde{G}$ respectively denote the mirror filters of $H$ and $G$ (i.e. $\tilde{h}(n) = h(-n)$ and $\tilde{g}(n) = g(-n)$).

It can then be shown [1] (cf. figure 2) that:

- ❖ $A_{2^j}I$ can be computed by convoluting $A_{2^{j+1}}I$ with $\tilde{H}$ in both dimensions and by sub-sampling by a factor of two in both dimensions ;
- ❖ $D^1_{2^j}I$ can be computed by :
  1. convoluting $A_{2^{j+1}}I$ with $\tilde{H}$ along the direction $y$ and by sub-sampling by a factor of two along this same direction ;
  2. convoluting the result of the step 1) with $\tilde{G}$ along the direction $x$ and by sub-sampling by a factor of two along this same direction.
- ❖ $D^2_{2^j}I$ may be computed by :
  1. convoluting $A_{2^{j+1}}I$ with $\tilde{G}$ along the direction $y$ and by sub-sampling by a factor of two along this same direction;
  2. convoluting the result of the step 1) with $\tilde{H}$ and along the direction $x$ and by sub-sampling by a factor of two along this same direction.
- ❖ $D^3_{2^j}I$ may be computed by :
  1. convoluting $A_{2^{j+1}}I$ avec $\tilde{G}$ with $\tilde{G}$ along the direction $y$ and by sub-sampling by a factor of two along this same direction;
  2. convoluting the result of the step 1) with $\tilde{G}$ along the direction $x$ and by sub-sampling by a factor of two along this same direction.

### *5.2 Construction of the tree structure with wavelet coefficients*

Once the wavelet transformation has been made up to the resolution $2^r (r \leq -1)$, we have available:

- ❖ an approximate image $A_{2^r}I$ ;
- ❖ Three detail images $D^1_{2^j}I$, $D^2_{2^j}I$, $D^3_{2^j}I$ per level of resolution $2^j$ with j=-1,...,r.

A tree structure of wavelet coefficients is then built by the zerotree technique [4]. The trees are built as follows (cf.figure 3) :

- Each pixel *p(x,y)* of the image $A_{2^r}I$ is the root of a tree ;
- Each root *p(x,y)* is assigned three offspring nodes designated by the wavelet coefficients of the three detail images $D^s_{2^r}I$ (s=1,2,3) localized at the same place *(x,y)* ;
- Owing to the sub-sampling by a factor of two performed by the wavelet transformation at each change in resolution, each wavelet coefficient $\alpha^s_{2^r}(x,y)$ (s=1,2,3) corresponds to a zone sized $2\times 2$ pixels in the detail image corresponding to the resolution $2^{r+1}$. This zone is localized at $(2x,2y)$ and all the wavelet coefficients belonging to it become the offspring nodes of $\alpha^s_{2^r}(x,y)$.

Recursively, the tree structure is constructed wherein each wavelet coefficient $\alpha^s_{2^u}(x,y)$ (s=1,2,3 and 0>u>r) possesses four offspring nodes designated by wavelet coefficients of the image $D^s_{2^{u+1}}I$ localized in the region situated in $(2x,2y)$ and sized $2\times 2$ pixels.

Once the tree structure is constructed, each wavelet coefficient $\alpha^s_{2^r}(x,y)$ (s=1,2, 3) corresponds to a region sized $2^{-r}\times 2^{-r}$ pixels in the detail image $D^s_{2^{-1}}I$.

### *5.3 Construction of the salience maps*

Starting from the tree structure obtained by the preceding step, we propose to build a set of–r salience maps (i.e. one salience map per level of resolution). Each salience map $S_{2^j}$ (j=-1,…,r) reflects the importance of the wavelet coefficients present at the corresponding resolution $2^j$. Thus, the more a wavelet coefficient will be deemed to be important with respect to the information that it conveys, the greater will be its salience value.

It must be noted that each wavelet coefficient gives preference to one direction (horizontal, vertical or oblique) depending on the detail image to which it belongs. However, we have chosen to favor no particular direction and have therefore merged the information contained in the three wavelet coefficients

$\alpha^1_{2^j}(x,y), \alpha^2_{2^j}(x,y), \alpha^3_{2^j}(x,y)$ whatever the level of resolution $2^j$ and whatever the localization *(x,y)* with $0 \le x < 2^{k+j}$ and $0 \le y < 2^{l+j}$.
Each salience map $S_{2^j}$ is sized $2^{k+j} \times 2^{l+j}$.

Furthermore, the salience of each coefficient with the resolution $2^j$ must take account of the salience of its offspring in the tree structure of the coefficients.

In order to take account of all these properties, the salience of a coefficient localized at *(x,y)* with the resolution $2^j$ is given by the following recursive relationship:

$$\begin{cases} S_{2^{-1}}(x,y) = \alpha_{-1}\left(\dfrac{1}{3}\displaystyle\sum_{u=1}^{3}\dfrac{D^u_{2^{-1}}(x,y)}{Max(D^u_{2^{-1}})}\right) \\ S_{2^j}(x,y) = \dfrac{1}{2}\left(\alpha_j\left(\dfrac{1}{3}\displaystyle\sum_{u=1}^{3}\dfrac{D^u_{2^j}(x,y)}{Max(D^u_{2^j})}\right) + \dfrac{1}{4}\displaystyle\sum_{u=0}^{1}\sum_{v=0}^{1} S_{2^{j+1}}(2x+u, 2y+v)\right) \end{cases}$$

**Equation 1: expression of the salience of a coefficient**

Where:

- ❖ $Max(D^s_{2^j})$ (s=1,2,3) denotes the maximum value of the wavelet coefficients in the detail image $D^s_{2^j}I$ ;
- ❖ $\alpha_k$ $(0 \le \alpha_k \le 1))$ is used to set the size of the salience coefficients according to the resolution level. It must be noted that we have $\sum_k \alpha_k = 1$.
- ❖ It must be noted that the salience values are standardized i.e. $0 \le S_{2^j}(x,y) \le 1$.

As can be seen in the Equation 1, the salience of a coefficient is a linear relationship of the wavelet coefficients. Indeed, as mentioned in section 4, we consider the salient points to be pixels of the image belonging to high-frequency regions. Now, a high wavelet coefficient $\alpha^s_{2^j}(x,y)$ (s=1,2,3) at the resolution $2^j$ denotes a high-frequency zone in the image $A_{2^{j+1}}I$ with the localization *(2x,2y)*. Indeed, the detail images are obtained by a high-pass filtering of the image $A_{2^{j+1}}I$ , each contour of $A_{2^{j+1}}I$ generates an elevated wavelet coefficient in one of the detail images with the resolution $2^j$, this coefficient corresponding to the

orientation of the contour.

Thus, the formulation of the salience of a given image in the Equation 1 is warranted.

### *5.4 Choice of the salient points*

Once the construction of the salience maps is completed, we propose a method in order to choose the most salient points in the original image.

To do this, we build a tree structure of the salience values from the –r built-up salience maps. In a manner similar to the building of the tree structure of the wavelet coefficients, we can build $2^{k+l+2r}$ trees of salience coefficients, each having a coefficient of $S_{2^r}$ as its root. As in the case of the zerotree technique, each of these coefficients corresponds to a zone sized 2x2 coefficients in the card $S_{2^{r+1}}$. It is then possible to recursively construct the tree in which each node is assigned four offspring in the salience map having immediately higher resolution. Figure 4 illustrates this construction.

In order to localize the most salient points in *I*, we carry out:

1. A descending-order sorting of the $2^{k+l+2r}$ salience values present in $S_{2^r}$ ;
2. Te selection of the maximum salience branch of each of the $2^{k+l+2r}$ trees thus sorted out.

In order to select this branch, it is proposed to scan the tree from the root. During this scan a selection is made, at each level of the tree, of the offspring node having the greatest salience value (cf. figure 5). We thus obtain a list of -r salience values:

$$SalientBranch=\{s_{2^r}(x_1,y_1),s_{2^{r+1}}(x_2,y_2),\cdots,s_{2^{-1}}(x_{-r},y_{-r})\}$$

with $\quad (x_k,y_k)=Arg\ Max\{s_{2^{r+(k-2)}}(2x_{k-1}+u,2y_{k-1}+v),0\leq u\leq 1,0\leq v\leq 1\}.$

From the most salient branches of each tree, the pixel of *I* chosen as being the most representative pixel of the branch is localized at $(2x_{-r},2y_{-r})$.

In practice, only a subset of the $2_{k+l+2r}$ trees is scanned. Indeed, for many applications, a search is made for a fixed number *n* of salient points. In this case, it is appropriate to scan only the *n* trees having the most salient roots.

## 6. Detailed description of at least one particular embodiment

In this section, we use the technical elements presented in the previous section for which we set the necessary parameters in order to describe a particular embodiment.

### *6.1 Choice of wavelet transformation*

As mentioned in section 5.1, we must first of all choose a wavelet base and a minimum resolution level $2^r\ (r \leq -1)$.

For this particular embodiment, we propose to use the Haar base and r = -4.

The Haar base is defined by:

$$\phi(x) = \begin{cases} 1 \text{ if } 0 \leq x < 1 \\ \text{else } 0 \end{cases}$$

for the scale function, and by:

$$\psi(x) = \begin{cases} 1 \text{ if } 0 \leq x < \frac{1}{2} \\ -1 \text{ if } \frac{1}{2} \leq x < 1 \\ \text{else } 0 \end{cases}$$

for the wavelet function.

### *6.2 Construction of the tree structure of the wavelet coefficients*

In this step, no parameter whatsoever is required. The process is therefore compliant with what is described in section 5.1.

### ***6.3 Construction of the salience maps***

In this step, we must choose the parameters $\alpha_k\ (-1 \geq k \geq r)$ used to adjust the importance given to the salience coefficients according to the level of resolution to which they belong.

In this particular embodiment, we propose to use $\alpha_k = \frac{-1}{r}\ \forall k \in [r,-1]$.

### *6.4 Choice of the salient points*

This step requires no parameter. The process is therefore compliant with what is described in section 5.4.

### *6.5 Experimental results*

The results obtained on natural images by using the parameters proposed in this particular embodiment are illustrated in figure 6.

### *6.6 Example of application*

Among the potential applications listed in the section 4, this section presents the use of salient points for the indexing of images fixed by the content.

#### *6.6.1 Purpose of image indexing*

Image indexing by content enables the retrieval, from an image database, of a set of images visually similar to a given image called a *request image*. To do this, visual characteristics (also called *descriptors*) are extracted from the images and form the signature of the image.

The signatures of the images belonging to the database are computed off-line and are stored in the database. When the user frequently submits a request image to the indexing engine, the engine computes the signature of the request image and cross-checks this signature with the pre-computed signatures of the database.

This cross-checking is made by computing the distance between the signature of the request image and the signatures of the database. The images most similar to the request image are then those whose signature minimizes the computed distance. Figure 7 illustrates this method.

The difficulty of image indexing then lies entirely in determining descriptors and robust distances.

#### *6.6.2 Descriptors based on the salient points of an image*

In this section, we propose to compute the signature of an image from a fixed number of salient points. This approach draws inspiration from [9].

A colorimetrical descriptor and texture descriptor are extracted at the vicinity of each of the salient points. The colorimetrical descriptor is constituted

by the 0 order (mean), $1^{st}$ order (variance) and $2^{nd}$ order moments in a neighborhood sized 3x3 around each salient point. The texture descriptor is constituted by the Gabor moments in a neighborhood sized 9x9.

Once the signature of the request image $R$ has been computed, the distance $D(R,I_j)$ between this signature and the signature of the $j^{th}$ image $I_j$ in the database is defined by:

$$D(R,I_j) = \sum_i W_i S_j(f_i),\ j = 1,\dots,N$$

where $N$ denotes the number of images in the database and $S_j(f_i)$ is defined by:

$$S_j(f_i) = (x_i - q_i)^T (x_i - q_i)$$

where $x_i$ and $q_i$ respectively designate the $i^{th}$ descriptor (for example i=1 for the colorimetrical descriptor and i=2 for the texture descriptor) of the $j^{th}$ image of the base and of the request $R$. The weights $W_i$ make it possible to modulate the importance of the descriptors relative to each other.

## Appendix A: an overview of the theory of wavelets

### *A.1 Introduction*

Wavelet theory [1][2][3] enables the approximation of a function (a curve, surface, etc.) at different resolution levels. Thus, this theory enables a function to be described in the form of a coarse approximation and of a series of details enabling the perfect reconstruction of the original function.

Such a multi-resolution representation [1] of a function therefore enables the hierarchical interpretation of the information contained in the function. To do this, the information is reorganized into a set of details appearing at different resolution levels. Starting from a sequence of resolution levels in ascending order $(r_j)_{j\in Z}$, the details of a function at the resolution level $r_j$ are defined as the difference of information between its approximation at the resolution $r_j$ and its approximation at the resolution $r_{j+1}$.

### *A.2 Notation*

Before presenting the bases of multi-resolution analysis in greater detail, in this section we shall present the notation that will be used in the document.

- The sets of integers and real numbers are respectively referenced *Z* and *R*.
- $L^2(R)$ denotes the vector space of the measurable and integrable 1D functions *f(x)*.
- For $f(x)\in L^2(R)$ and $g(x)\in L^2(R)$, the scalar product of *f(x)* and *g(x)* is defined by:

$$\langle f(x),g(x)\rangle=\int_{-\infty}^{+\infty}f(u)g(u)du\,.$$

- For $f(x)\in L^2(R)$ et $g(x)\in L^2(R)$, the convolution of *f(x)* and *g(x)* is defined by:

$$f*g(x)=\int_{-\infty}^{+\infty}f(u)g(x-u)du\,.$$

- $L^2(R^2)$ denotes the vector space of the functions *f(x,y)* of two measurable and integrable variables.
- For $f(x,y)\in L^2(R^2)$ and $g(x,y)\in L^2(R^2)$, the scalar

product of $f(x,y)$ and $g(x,y)$ is defined by:

$$\langle f(x,y), g(x,y)\rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(u,v)g(u,v)dudv.$$

### A.2 Properties of multi-resolution analysis

This section intuitively presents the desired properties of the operator enabling the multi-resolution analysis of a function. These properties come from [1].

Let $A_{2^j}$ be the operator that approximates a function $f(x) \in L^2(R)$ with the resolution $2^j$ $(j \geq 0)$ (i.e. $f(x)$ is defined by $2^j$ samples).

The following are the properties expected from $A_{2^j}$ :

1. $A_{2^j}$ is a linear operator. If $A_{2^j} f(x)$ represents the approximation of $f(x)$ with the resolution $2^j$, then $A_{2^j} f(x)$ should not be modified when it is again approximated at the resolution $2^j$. This principle is written as $A_{2^j} \circ A_{2^j} = A_{2^j}$ and shows that the operator $A_{2^j}$ is a projection vector in a vector space $V_{2^j} \subset L^2(R)$. This vector space may be interpreted as the set of all the possible approximations at the resolution $2^j$ of the functions of $L^2(R)$.

2. Among all the possible approximations of $f(x)$ with the resolution $2^j$, $A_{2^j} f(x)$ is the most similar to $f(x)$. The operator $A_{2^j}$ is therefore an orthogonal projection on $V_{2^j}$.

3. The approximation of a function at the resolution $2^{j+1}$ contains all the information necessary to compute the same function at the lower resolution $2^j$. This property of causality induces the following relationship:

$$\forall j \in Z, V_{2^j} \subset V_{2^{j+1}}.$$

4. The operation of approximation is the same at all values of resolution. The spaces of the approximation function may be derived from one another by a change of scale corresponding to the difference of resolution.

$$\forall j \in Z, f(x) \in V_{2^j} \Leftrightarrow f(2x) \in V_{2^{j+1}}.$$

5. When an approximation of *f(x)* at the resolution $2^j$, is computed, a part of the information contained in *f(x)* is lost. However, when the resolution tends toward infinity, the approximate function must converge on the original function *f(x).* In the same way, when the resolution tends toward zero, the approximate function contains less information and must converge on zero.

Any vector space $(V_{2^j})_{j\in Z}$ that complies with all these properties is called the multi-resolution approximation *de* $L^2(R)$.

## A.3 Multi-resolution analysis of a 1D function

### A.3.1 Search for a base of $V_{2^j}$

We have seen in section A.2 that the approximation operator $A_{2^j}$ is an orthogonal projection in the vector space $V_{2^j}$. In order to numerically characterize this operator, we must find an orthonormal base of $V_{2^j}$.

$V_{2^j}$ being a vector space containing the approximations of functions of $L^2(R)$ with the resolution $2^j$, any function $f(x) \in V_{2^j}$ may be seen as a vector with $2^j$ components. We must therefore find $2^j$ base functions.

One of the main theorems of the theory of wavelets stipulates that there is a single function $\Phi(x) \in L^2(R)$, called a scale function, from which it is possible to define $2^j$ base functions $\Phi_i^j(x)$ de $V_{2^j}$ by expansion and translation of $\Phi(x)$ :

$$\Phi_i^j(x) = \Phi(2^j x - i), i = 0, \cdots, 2^j - 1.$$

Approximating a function $f(x) \in L^2(R)$ at the resolution $2^j$ therefore amounts to making an orthogonal projection *f(x)* on the $2^j$ basic functions $\Phi_i^j(x)$. This operation consists in computing the scalar product of *f(x)* with each of the $2^j$ basic functions $\Phi_i^j(x)$ :

$$\begin{aligned} A_{2^j} f(x) &= \sum_{k=0}^{k=2^j-1} \langle f(u), \Phi_k^j(u) \rangle \Phi_k^j(x) \\ &= \sum_{k=0}^{k=2^j-1} \langle f(u), \Phi(2^j u - k) \rangle \Phi(2^j u - k). \end{aligned}$$

It can be shown [1] that $A_{2^j} f(x)$ may be reduced to the convolution of

$f(x)$ with the low-pass filter $\Phi(x)$, assessed at the point $k$:

$$A_{2^j}f = (f(u) * \Phi(-2^j u))(k), k \in Z.$$

Since $\Phi(x)$ is a low-pass filter, $A_{2^j}f$ may be interpreted as a low-pass filtering followed by a uniform sub-sampling.

### *A.3.2 Construction of the multi-resolution analysis*

In practice, the functions $f$ to be approximated (signal, image, etc.) are discrete. Let it be assumed that the original function $f(x)$ is defined on $n = 2^k\ (k \in Z)$ samples. The maximum resolution of $f(x)$ is then $n$.

Let $A_n f$ be the discrete approximation of $f(x)$ at the resolution level $n$. According to the property of causality, it is claimed (cf. section A.2) that $A_{2^j}f$ can be computed from $A_n f$ for every value of $j < k$.

Indeed, in computing the projection of the $2^j$ basic functions $\Phi_i^j(x)$ of $V_{2^j}$ on $V_{2^{j+1}}$, it can be shown that $A_{2^j}f$ can be obtained by convoluting $A_{2^{j+1}}f$ with the low-pass filter corresponding to the scale function and by sub-sampling the result by a factor of 2:

$$A_{2^j}f(u) = \sum_{k=0}^{2^{j+1}-1} h(k-2u) A_{2^{j+1}}f(k), 0 \le u < 2^j - 1$$

with $h(n) = \langle \Phi(2u), \Phi(u-n) \rangle, \forall n \in Z$.

### *A.3.3 The detail function*

As mentioned in the property (5) of section A.3, the operation which consists in approximating a function $f(x)$ at the resolution $2^j$ on the basis of an approximation at the resolution $2^{j+1}$ causes a loss of information.

This loss of information is contained in a function called a detail function at resolution level $2^j$ and referenced $D_{2^j}f$. It must be noted that knowledge of $D_{2^j}f$ and $A_{2^j}f$ enables the perfect reconstruction of the approximate function $A_{2^{j+1}}f$.

The detail function at the resolution level $2^j$ is obtained by projecting the original function $f(x)$ orthogonally on the orthogonal complement of $V_{2^j}$ in $V_{2^{j+1}}$. Let $W_{2^j}$ be this vector space.

To calculate this projection numerically, we need to find an orthonormal base of $W_{2^j}$ i.e. $2^j$ base functions. Another important theorem of the wavelet theory stipulates that, through a scale function $\Phi(x)$, it is possible to define $2^j$ base functions of $W_{2^j}$. These base functions $\Psi_i^j(x)$ are obtained by expansion and translation of a function $\Psi(x)$ called a wavelet function:

$$\Psi_i^j(x) = \Psi(2^j x - i), i = 0, \cdots, 2^j - 1.$$

In the same way as for the construction of the approximation $A_{2^j} f$, it can be shown that $D_{2^j} f$ can be obtained by a convolution of the original function *f(x)* with the high-pass filter $\Psi(x)$ followed by a sub-sampling by a factor of $2^j$ :

$$D_{2^j} f = (f(u) * \Psi(-2^j u))(k), k \in Z.$$

### *A.4.5 Extension to the multi-resolution analysis of 2D functions*

This section presents the manner of extending multi-resolution analysis by wavelets to the functions of $L^2(R^2)$ such as images.

This is done by using the same theorems as the ones used earlier. Thus if $V_{2^j}$ denotes the vector space of the approximations of $L^2(R^2)$ at the resolution $2^j$, it can be shown that it is possible to find an orthonormal base of $V_{2^j}$ by expanding and translating a scale function $\Phi(x,y) \in L^2(R^2)$ :

$$\Phi_i^j(x,y) = \Phi(2^j x - i, 2^j y - j), (i,j) \in Z^2.$$

In the particular case of the separable approximations of $L^2(R^2)$, we have $\Phi(x,y) = \Phi(x)\Phi(y)$ where $\Phi(x)$ is a scale function of $L^2(R)$. In this case, the multi-resolution analysis of a function of $L^2(R^2)$ is done by the sequential and separable processing of each of the dimensions $x$ and $y$.

As in the 1D case, the detail function at the resolution $2^j$ is obtained by an orthogonal projection of *f(x,y)* on the complement of $V_{2^j}$ in $V_{2^{j+1}}$, written as $W_{2^j}$. In the 2D case, it can be shown that if $\Psi(x)$ denotes the wavelet function associated with the scale function $\Phi(x)$, then the three functions defined by:

$$\Psi^1(x,y) = \Phi(x)\Psi(y)$$
$$\Psi^2(x,y) = \Psi(x)\Phi(y)$$
$$\Psi^3(x,y) = \Psi(x)\Psi(y)$$

are wavelet functions of $L^2(R^2)$. Expanding and translating these three wavelet functions gives an orthonormal base of $W_{2^j}$ :

$$\Psi_j^1(x,y) = \Phi\Psi(2^j x - k, 2^j y - l)$$
$$\Psi^2(x,y) = \Psi\Phi(2^j x - k, 2^j y - l)$$
$$\Psi^3(x,y) = \Psi\Psi(2^j x - k, 2^j y - l).$$

The projection of f(x,y) on these three base functions of the base of $W_{2^j}$ gives three detail functions:

$$D_{2^j}^1 f = f(x,y) * \Phi^j(-x)\Psi_j(-y)$$
$$D_{2^j}^2 f = f(x,y) * \Psi^j(-x)\Phi_j(-y)$$
$$D_{2^j}^1 f = f(x,y) * \Psi^j(-x)\Psi_j(-y)$$

## Appendix B : References

[1] Mallat S., "*A Theory for Multiresolution Signal Decomposition: the Wavelet Representation*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, July 1989, pp. 674-693.

[2] Stollnitz E.J., DeRose T.D., Salesin D., "*Wavelets for Computer Graphics: A Primer-Part 1*", IEEE Computer Graphics and Applications, Mai 1995, pp.76-84.

[3] Stollnitz E.J., DeRose T.D., Salesin D., "*Wavelets for Computer Graphics: A Primer-Part 2*", IEEE Computer Graphics and Applications, July 1995, pp.75-85.

[4] Shapiro J.M., " *Embedded Image Coding Using zerotrees of Wavelet Coefficients* ", IEEE Transactions on Signal Processing, Vol. 41, No. 12, December 1993, pp. 3445-3462.

[5] Schmid C., Mohr R. and Bauckhage C., " *Evaluation of Interest Point Detectors*", International Journal of Computer Vision, Vol. 37, No 2, pp. 151-172, 2000.

[6] Gouet V. and Boujemaa N., "About Optimal Use of Color Points of Interest for Content-Based Image Retrieval", INRIA research report, No 4439, April 2002.

[7] Harris C. and Stephens M., "*A Combined Corner and Edge Detector*", Proceedings of the 4th Alvey Vision Conference, 1988.

[9] Sebe N. and Lew M.S., "*Salient Points for Content-based Retrieval*", Proceedings of British Machine Vision Conference, Manchester, 2001.

[10] Bres S. and Jolion J.M., "*Detection of Interest Points for Image Indexation*".

[11] Loupias E. and Sebe N., "*Wavelet-based Salient Points for Image Retrieval* ", Research Report RR 99.11, INSA Lyon, 1999.